DRAFT – PLEASE DO NOT CITE. COMMENTS ARE WELCOME

***

Thought Experiments in Biology

13 October 2014

Guillaume Schlaepfer and Marcel Weber

1. Introduction

Unlike in physics, the category of thought experiment is not very common in biology. At least there are no classic examples that are as important and as well-known as the most famous thought experiments in physics, such as Galileo's, Maxwell's or Einstein's. The reasons for this are far from obvious; maybe it has to do with the fact that modern biology for the most part sees itself as a thoroughly empirical discipline that engages either in real natural history or in experimenting on real organisms rather than fictive ones. While theoretical biology does exist and is recognized as part of biology, its role within biology appears to be more marginal than the role of theoretical physics within physics. It could be that this marginality of theory also affects thought experiments as sources of theoretical knowledge.

Of course, none of this provides a sufficient reason for thinking that thought experiments are really unimportant in biology. It is quite possible that the common perception of this matter is wrong and that there are important theoretical considerations in biology, past or present, that deserve the title of thought experiment just as much as the standard examples from physics. Some such considerations may even be widely known and considered to be important, but were not recognized as thought experiments. In fact, as we shall see, there are reasons for thinking that what is arguably the single

most important biological work ever, Charles Darwin's *On the Origin of Species*, contains a number of thought experiments. There are also more recent examples both in evolutionary and non-evolutionary biology, as we will show.

Part of the problem in identifying positive examples in the history of biology is the lack of agreement as to what exactly a thought experiment is. Even worse, there may not be more than a family resemblance that unifies this epistemic category. We take it that classical thought experiments show the following characteristics: (1) They serve directly or indirectly in the *non-empirical epistemic evaluation* of theoretical propositions, explanations or hypotheses. (2) Thought experiments somehow appeal to the *imagination* (so a purely mechanical deduction or calculation from theoretical principles would not count as a thought experiment). (3) They involve *hypothetical scenarios*, which may or may not be fictive. In other words, thought experiments suppose that certain states of affairs hold (irrespectively of whether they obtain in the actual world or not) and then try to intuit what would happen in a world where these suppositions are true.

We want to examine in the following sections if there are episodes in the history of biology that satisfy these criteria. As we will show, there are a few episodes that might satisfy all three of these criteria, and many more if the imagination criterion (2) is dropped or understood in a lose sense. In any case, this criterion is somewhat vague in the first place, unless a specific account of the imagination is presupposed. There will also be issues as to what exactly "non-empirical" means. In general, for the sake of discussion we propose to understand the term "thought experiment" here in a broad rather than a narrow sense here. We would rather be guilty of having too wide a conception of thought experiment than of missing a whole range of really interesting examples.

2. Darwin

There are several places in the work of Darwin where he resorts to hypothetical scenarios in the course of presenting his theories. The best known cases are those from Darwin's *On the Origin of Species* (1859) discussed by Lennox (1991), but similar imaginary cases can be found in *The Descent of Man* (1871) as well as in his essay of 1844 and it is likely that there are even more. The reason is that Darwin, due to the speculative character of his endeavor, very often presents hypothetical explanations and each of them could potentially qualify as thought experiment. In any case, we will have to restrict ourselves here to the presentation of the canonical cases discovered by Lennox and discuss their epistemic role in the context of the *Origin* as well as their specific features.

Lennox argues that Darwin's hypothetical cases play a crucial role in his argumentation in the *Origin*. According to Lennox, their role is to show the explanatory potential of Darwin's theory rather than to serve as evidence for the theory's truth. Thus, they aim at showing that the theory of natural selection *can* explain such things as species transformation, appearance of new species, and the extraordinary adaptation of organisms to their environment, but not that it actually *does* explain these phenomena. The following "imaginary illustrations", as Darwin calls them, probably prompted the well-known criticism by Fleeming Jenkin (Hull 1973), in which Jenkin also uses hypothetical scenarios against Darwin's idea of natural selection.[1] Jenkin thus provided

[1] Although Jenkin falsely concludes from his imaginary cases that natural selection cannot explains speciation, his criticism goes right to the point by targeting the importance of the

what Brown (1991) calls a *counter thought experiment*, which led Darwin to refine his original thought experiment.

The following case is one of the two "imaginary illustrations" presented by Darwin in the end of the fourth chapter of the Origin. It is the paradigm of Lennox's "Darwinian thought experiments":[2]

> Let us take the case of a wolf, which preys on various animals, securing some by craft, some by strength, and some by fleetness; and let us suppose that the fleetest prey, a deer for instance, had from any change in the country increased in numbers, or that other prey had decreased in numbers, during that season of the year when the wolf is hardest pressed for food. I can under such circumstances see no reason to doubt that the swiftest and slimmest wolves would have the best chance of surviving, and so be preserved or selected. […]
>
> Now, if any slight innate change of habit or of structure benefited an individual wolf, it would have the best chance of surviving and of leaving offspring. Some of its young would probably inherit the same habits or structure, and by the repetition of this process, a new variety might be formed which would either supplant or coexist with the parent-form of wolf. (Darwin, 1859, pp. 90–91)

---

frequency of variation and the problem of inheritance type, which led Darwin to slightly modify his thought experiments in the later versions of the *Origin* (Lennox 1991, Morris 1994, Bulmer 1999).

[2] This thought experiment is directly followed by another one, which we will not reproduce here for the sake of brevity. It pertains to the explanation of mutual adaptations of bees and flowers. Other thought experiments can be found in Chapters Six, on the evolution of complex organs and Seven, on the evolution of instincts.

Lennox considers this passage as a thought experiment, for it displays an imaginary situation and provides support to Darwin's theory. Furthermore, he points out three aspects of Darwin's illustrations, which, he claims, constitute jointly sufficient criteria for characterizing a thought experiment: First, the thought experiment should evoke concrete objects or processes in order to "give […] the feeling of experimentation", which cannot be obtained by staying at the level of abstract theories. Second, the described situation should be plausible. This is normally ensured by referring to familiar objects and processes that "could happen, in a fairly robust sense of 'could'". Finally, it should relate to the theoretical claim to which it lends support by instantiating some of its essential features (see Lennox, 1991, pp. 229–230).

Concerning the epistemic role of these thought experiments, Lennox's thesis that Darwin did not intend them as providing direct evidence for natural selection is supported in several ways. Darwin's use of the expression "imaginary illustrations" as well as the subjunctive mood already give quite strong indications. Furthermore, Darwin makes explicit that the aim of the chapter is not to provide a proof for the theory of natural selection: "[w]hether natural selection has really thus acted in nature […] must be judged of by the general tenour and balance of evidence given in the following chapters" (Darwin, 1859, p. 127). On the other hand, the claim according to which these thought experiments aim at showing the explanatory power of the theory also fits the general plan of the *Origin*. They are located in Chapter Four, which, according to a synoptic view of 'Darwin's argument' proposed by Waters (2003), has the role of showing that natural selection is an "adequate" explanation in the Herschelian sense of establishing a *vera causa*, which precisely amounts to showing that the theory could explain speciation and adaptation.

The view that the role of thought experiments is to evaluate a theory's explanatory potential is shared also by Kuhn (1977). But according to Kuhn, thought experiments can advance science only by showing a theory's failure. According to Lennox, there is no such asymmetry. Thought experiments, like in the case presented above, can also provide positive support regarding a theory's explanatory potential. This probably constitutes the most specific feature of Lennox's concept of "Darwinian thought experiments".

Restricting the role of thought experiments to an assessment of the potential or possible truth of an explanation grants them a legitimate role in science even if that role is "independent of empirical support of a theory's truth" (Lennox, 1991, p. 237). To illustrate this, Lennox takes the case of the philosophical debate over adaptationism (Gould & Lewontin, 1979, Kitcher, 1982, p. 60, Kitcher, 1985, p. 226, cited by Lennox, 1991, p. 240). This debate is rooted in a criticism of the tendency to explain everything in nature with help of natural selection, which amounts to considering any feature of an organism as an adaptation. On Lennox's view, adaptationism helps to explore natural selection theory's explanatory scope no matter whether it (adaptationism) is true or false. Similarly, thought experiments do not speak for the truth or falsity of a theory, but they can serve to fathom and bring support to a theory's explanatory potential.

Lennox's account of Darwinian thought experiments raises many interesting issues. A first issue concerns the identification and the limits of the epistemic category of thought experiments, another one the psychological aspects related to the evaluation of the explanatory power of a theory by means of Darwinian thought experiments. We shall discuss these two issues in turn.

Concerning the first issue, it should be noted that, although Darwin calls them "imaginary illustrations", the thought experiments of Chapter Four in the *Origin* are

supported by observational data. For instance, in the wolf pack case, Darwin appeals to evidence regarding the heritability of fleetness observed in greyhounds or of the tendency to prey on specific animals observed in domestic cats. He also mentions two varieties of wolves observed in the Catskill Mountains in New York that may instantiate the process proposed in the thought experiment. Lennox interprets this appeal to experimental data as a way to improve the plausibility of the thought experiments, but it also restricts the extent to which these illustrations can be considered as imaginary. The imaginary part of the illustration here seems to be the explanatory process rather than the particular case at stake, which is real. The same can be said of other Darwinian thought experiments in the *Origin*.

Should any hypothetical scenario pertaining to a theoretical claim about a particular matter of fact considered to be a thought experiment? If not, then what restriction could prevent such an explosion of the category? Kuhn (1977), on his part, provides a restriction by limiting thought experiments to situations of conceptual conflict arising during scientific crisis. But why should this be so? By generalizing Kuhn's definition to imagined cases directly supporting a theory, thought experimenting appears to be a much broader category than is usually supposed. The only distinction between a thought experiment and any theoretical claim submitted to the diligent judgment of the reader seems to be, in Lennox's view, the appeal to concrete objects or processes.

This brings us to the second issue, the psychological aspects of thought experimenting and the requirement of concreteness. To what extend and how does the appeal to concrete object or processes help in evaluating the explanatory potential of a theory? It should be noted that one of the main differences between the mind and the laboratory is that the objects manipulated within the mind are abstract to begin with, therefore it isn't obvious why thought experiments should be restricted to imagining

concrete cases. A possible argument might be that supporting a theory must necessarily involve a reference to one of its particular instances, as suggested by Lennox. But even if such a principle is granted for *empirical* support, it is not clear if this principle transfers to *thought* experiments, which often seem to appeal to a *type* of instance, not any concrete case (cf. Maxwell's demon-type scenarios, which seem to be generic rather than particular).

3. Population Genetics and Natural Selection Theory

Darwin's theory of natural selection was thoroughly transformed in the early 20th Century due to the work of mathematical population geneticists, most notably R.A. Fisher, J.B.S. Haldane and S. Wright (Provine 1971). These scientists provided a series of mathematical models describing the evolution of gene frequencies in populations of sexually reproducing organisms under the influence of evolutionary forces such as natural selection or random drift. This work culminated in a synthesis of Darwinian thought with more recent discoveries in genetics, in particular Mendel's laws as well as the chromosomal mechanisms of inheritance. Before these mathematical models existed, the theory of natural selection was still controversial and thought to be incompatible with modern genetics. The reason was that genetics was seen as being concerned only with discrete mutations, which seemed to suggest a saltationist theory of evolution, i.e., evolution through discontinuous leaps or "hopeful monsters" rather than Darwin's idea of a gradual adaptive process (Mayr 1982). But especially the work of Fisher proved that Mendelian inheritance can give rise to continuous variation that provides the material for natural selection to act on, and that selection is theoretically effective in changing the frequencies of genes in a population. In addition, Fisher

provided theoretical models that could explain the abundance of species that reproduce sexually as well as other frequently encountered traits.

The introduction to Fisher's extremely influential book *The Genetical Theory of Natural Selection* (Fisher 1930) appears to announce a thought experiment:

> No practical biologist interested in sexual reproduction would be led to work out the detailed consequences experienced by organisms having three or more sexes; yet what else should he do if he wishes to understand why sexes are in fact, always two? (Fisher 1930, ix).

In this passage, Fisher seems to suggest that, in order to understand why actual organisms normally come in two sexes, it is necessary to examine hypothetical organisms with three or more sexes. Only if the evolutionary consequences of having one, two, three or more sexes are understood, can we hope to understand why two sexes are so common. While this looks like a thought experiment, the curious reader will look for it in vain in the rest of the book. Fisher does not come back to the issue; he only offers a model describing the advantages of sexual reproduction in general over asexual reproduction. Furthermore, as Weisberg (2013) has pointed out, the problem as stated in Fisher's book is underspecified. For example, he doesn't say if all the 3 or *n* sexes are supposed to be necessary for producing offspring or if different mating types (i.e., subtypes of a species that can only produce offspring with members from a different type) would also count. Depending on how the 3- or *n*-sex biology is understood, it is not fictive at all but has well-known precedents in nature (see Weisberg 2013, pp. 131-134). However, Fisher's intention was clearly to describe a non-existent entity with his three or more sexes.

Perhaps Fisher's example should not be taken too literally; in fact, its purpose according to Fisher is only to illustrate the advantages of approaching evolutionary problems from a mathematical point of view, which Fisher claims, is characterized by a specific kind of imagination.[3] But Fisher's thought experiment also illustrates a general strategy that characterizes his approach and that is very often used in evolutionary biology: When considering a trait, biologists first define a character space that contains all the possible values that the trait can take. For example, take the ratio of male and female offspring that the female of a given species produces on average. In many species, this ratio is close to 1:1, but in theory it could take any value, so long as it is a rational number. Let us call this space of values the character space. This space defines a set of *logical* possibilities; thus the idea is not that each value of the space could actually be realized. As Darwin already noted, there could be "laws of growth" that prevent some character values of being realized. Thus, there will be a subspace of character values that is logically as well as biologically possible. Evolution will only be capable of producing organisms within the subspace of biologically possible character values.

This approach can be seen at work in a field known as "life history theory" (Stearns 1992). Life histories (or life cycles) involve traits such as the age of onset of

[3] Fisher suggests that "[…] the intelligence, properly speaking, is little influenced by the effects of training. What is profoundly susceptible of training is the imagination, and mathematicians and biologists seem to differ enormously in the manner in which their imaginations are employed" (Fisher 1930, p. viii). Furthermore, "[i]t seems impossible that full justice should be done to the subject in this way, until there is built up a tradition of mathematical work devoted to biological problems, comparable to the researches upon which a mathematical physicist can draw in the resolution of special difficulties" (Fisher 1930, p. x).

reproduction, the number of offspring produced, as well as the average life span or longevity. Variation in these traits is enormous; some organisms go through a long larval phase before they mature, just to reproduce once and then die quickly. Other species enjoy a steady output of offspring almost throughout their lifetime, some over hundreds of seasons. Life history theory tries to explain why each species has the life history traits it does. But in order to do so, it needs to consider the space of all logically possible trait combinations. Typically, life history theorists will try to show that an organism's life cycle is somehow adapted or optimized. But of course, evolution cannot produce any combination of life history traits. There are constraints or trade-offs that limit the range of possible values that the different life history characters can take. For example, it is widely thought by life history theorists that reproduction has a cost: Producing more offspring reduces longevity, both because it is inherently dangerous and because the metabolic resources used for reproduction are diverted from the maintenance of the body. Thus, it is not possible to maximize both the number of offspring and longevity at the same time.

The role of such constraints or trade-offs is sometimes explained with the help of a fictive creature known as a "Darwinian demon". This is an organism "which can optimize all aspects of fitness simultaneously" (Law 1979, p. 399). By an "aspect of fitness", Law means a character such the number of offspring or the life span. Obviously, a genotype that produces more offspring and lives longer than his conspecifics would quickly dominate the population, such that this genotype will be the only one present. His fitness would be vastly superior. So why don't naturally occurring organisms become fitter and fitter, eventually turning into Darwinian demons? This is one of the central questions of life history theory. There are many answers that have

been given to this question, but they usually involve some kind of trade-off or constraint that prevents a species of maximizing all aspects of fitness at the same time.

Thus, even if Fisher's example of a three- or *n*-sex biology is questionable, the principle that he was trying to illustrate is an important one: When giving an evolutionary explanation, biologists cannot be content to consider just actual organisms. They must also examine a range of logically possible organisms that are defined by a character space. The explanation will then show why the traits of actual organisms take certain values within this character space. The explanation will typically be based on the assumption that all other trait combinations have been eliminated by natural selection. But showing that some trait combination is actually the fittest that is available given the constraints, trade-offs etc. requires that the fitness values for different possible organisms be estimated. Of course, the same strategy is necessary in order to show that some trait has not been optimized by natural selection, so there is no presumption of adaptationism here (cf. Gould and Lewontin 1979).

Fisher's example aside, can we find in the explanatory practice that we have just outlined any use of thought experiment? What about the idea of a Darwinian demon, this "mythical entity […] that grows quickly, breeds fast, outcompetes all and never ages" (Bonsall 2006), p. 120)? Is this demon comparable to Maxwell's demon? In fact, there are interesting parallels in terms of the role that these two demons play. Maxwell imagined two chambers filled with gas of different temperature (see article on thought experiments in physics, this volume?). In between the two chambers there would be a door that is controlled by the demon. The demon would open the door whenever a faster than average molecule were about to pass from the cooler to the warmer gas, and also when a slower than average molecule was about to pass from the warmer to the cooler side. Thus, heat would spontaneously flow from the cooler to the warmer body, in

violation of the Second Law of thermodynamics. Maxwell's demon describes a scenario that is supposed to be *physically impossible* and challenges theoretical physics to give an explanation of what physical principles prevent Maxwell's imaginative scenario from being realized. (The standard answer is that Maxwell's demon would need to acquire information about the velocity and trajectory of gas molecules, which would generate enough entropy to offset the entropy reduction generated by the demon's gate controlling of molecules.) It could be suggested that the role of the Darwinian demon is similar: It describes a biologically impossible scenario such that biologists are challenged to explain why it is impossible. While Maxwell's demon scenario seems more imaginative, both are fictions and thus involve the imagination.

But it should also be noted that the Darwinian demon is hardly indispensable for evolutionary biologists to do their work (unlike, perhaps, some thought experiments in physics). This is evident because there was important theoretical work done in life history theory before the demon was introduced in 1979 (by Richard Law). Fisher's own work is a case in point. The evolutionary biologist Michael B. Bonsall refers to the Darwinian demon as an "iconic representation" that serves to "focus the thoughts and ideas presented" (p. 120). This would suggest a mainly pedagogical role for the Darwinian demon, perhaps much like Fisher's case of three sexes.

Are there any reasons for thinking that thought experimenting plays a more prominent role in population genetics, i.e., a role that goes beyond mere illustration? In fact, there could be such reasons. Sober (2011) has argued that some mathematical models of evolutionary biology provide *a priori* causal knowledge. This seems like quite a radical claim at first, as causal knowledge is widely thought to be empirical in nature (due to the extremely influential arguments by David Hume). However, Sober's claim is not that population genetic models such as Fisher's sex ratio model mentioned

above can provide any knowledge about *actual* causes. Rather, what these models can do is to tell the biologist that under certain conditions, conditions that may be real or hypothetical, some state of a population of organisms *would promote* some other state. For example, an unbalanced state of a population with respect to sex ratio *would promote* an increase, on an evolutionary time scale, in females that have a tendency for producing more offspring of the minority sex.

Sober's argument is philosophically sophisticated and we lack the space here to give it a proper treatment. Clearly, Sober's "would promote"-locution is in need of explication; it is not clear if it can be interpreted causally while at the same time maintaining it's *a priori* status (see Lange and Rosenberg 2011).

4. Molecular Biology

For the most part, discussion about thought experiments in biology has focused on evolutionary biology. However, there are also candidates to be found in other fields of biology. We shall discuss two examples from mid-20th century molecular biology. The first example concerns protein synthesis and the genetic code, the other is about protein folding and the so-called "Levinthal paradox".

In 1958, the co-discoverer of the double-helix structure of DNA Francis Crick published a remarkable paper simply titled "on protein synthesis" (Crick 1958). In this paper, he attempted to draw together all that was known at that time about the synthesis of proteins by living cells and then formulate the outlines of a mechanism. The paper became famous for containing a hypothesis that Crick dubbed the "Central Dogma of Molecular Biology" (it was really a hypothesis at this stage; Crick later reported that he misapplied the term "dogma"). According to the Central Dogma, genetic information flows from DNA to RNA to protein, but not in the other direction. What Crick meant by

this is that the sequence of DNA nucleotide bases determines the sequence of RNA bases, which in turn determines the sequence of amino acids in proteins. The converse is not true according to Crick; RNA does not determine DNA sequence (which turned out to be incorrect) and protein sequence does not determine RNA sequence (which is still considered to be correct today).

For Crick, this hypothesis was a solution to one of the great puzzles of biology: namely, the question of how a living cell can make thousands of different specific protein molecules. While much was known in 1958 how a cell can make other biomolecules (carbohydrates, lipids), the case of the proteins was trickier. The reason was the following: Other biomolecules are typically made by specific enzymes. But obviously, this cannot be how the cell can make different proteins because enzymes are proteins themselves. So to make a specific protein such as hemoglobin, the cell would need a specific enzyme. But to make this enzyme, it would need another enzyme and so on. This mechanism would thus generate an infinite regress. Crick concluded that some kind of code was necessary to specify the amino acid composition of all the different proteins made by a cell. According to this idea, the amino acid sequence of each protein was determined by the DNA base sequence of a gene. The code, which came to be known as the "genetic code" later, determines which amino acid sequence is specified by any arbitrary DNA sequence, just like the Morse code determines what letter of the alphabet is specified by a combination of short and long signals.

Crick thus not only predicted the existence of a genetic code, he also provided some constraints as to how this code might look like. (We are simplifying the history here. Crick was not the first to engage in such considerations, but his speculations turned out to be remarkably accurate. For historical details see Judson 1979). Since there are exactly 20 different naturally occurring amino acids that compose all the proteins in any

living organism, it was clear that more than one nucleotide base was necessary to encode these amino acids. For there are only four nucleotide bases, usually abbreviated as A, T, G and C. Let us image that each amino acid were specified by two bases in the genetic code. Because there are $2^4 = 16$ combinations of nucleotides, this would not be enough for 20 amino acids. Therefore, the code had to be at least a triplet code. But there are already $3^4 = 64$ triplet combinations of four bases, therefore the code had to be redundant. This means that at least some amino acids must correspond to several base triplets, or else there would have to be meaningless codons.

Remarkably, all this is exactly what was found in the 1960s through much painstaking experimental work using protein-synthesizing cell extracts that were programmed with artificial polynucleotides. The genetic code was "cracked" and the exact mapping of base triplets to amino acids was unraveled. But Crick (with some help from other scientists) had figured out some of this on the basis of theoretical considerations alone.

We suggest that some parts of Crick's reasoning show some of the characteristic marks of a thought experiment: In particular, Crick examined several hypothetical or counterfactual scenarios, including a scenario where each protein is made by a specific enzyme as well as a scenario with a duplet code. He then showed that these scenarios were not possible, thus lending support to other scenarios such as the triplet code. Thus, as in the other cases discussed so far, thought experimenting was used in order to explore logical spaces of possibility and to identify some regions in these spaces as not only logically but biologically possible. Having thus noticed an important similarity to classic thought experiments, we should also point out that some aspects of this example fit the category less well. In particular, can we say that Crick used his imagination when reasoning about the genetic code? The difficulty with this is that his considerations

appear to be quite abstract and based more on a principle of causality and simple math than on imagining a fictive scenario (such as in the case of Maxwell's demon). But all imagination might require principles and some of the classical thought experiments from physics are also quite abstract and need not imagine a particular scenario (e.g., Galileo's). Thus, it is difficult to say if this is a case of imagining or not.

Another issue is the exact epistemic role of Crick's considerations. Did it consist in the non-empirical evaluation of theoretical hypotheses? In a sense, one could say that Crick was examining the truth of certain hypotheses about the mechanism of protein synthesis. On the other hand, it is also clear that he did not consider these considerations as definitive proof of any theoretical hypothesis. The mechanism of protein synthesis was clearly to be established experimentally, which it eventually was. So perhaps one could say that what we have here is a different epistemic use of thought experiment, perhaps a use that is properly located in the context of discovery rather the context of justification (cf. Weber 2005, Ch. 3). In other words, the thought experiment was used here in order to *generate* rather than *evaluate* theoretical hypotheses.

Another example from the history of molecular biology is the so-called "Levinthal paradox". In 1969, Cyrus Levinthal surprised the scientific community with a simple argument that called into question the received view of how protein chains fold into their three-dimensional structure (Levinthal 1969) . According to the received view, a newly synthesized protein chain will randomly twist and move about until it has found the three-dimensional structure that corresponds to its lowest energy state. This state is given by the intramolecular interactions (hydrophobic interactions and hydrogen bonds) that are possible between the different parts of the molecule. Now, Levinthal made a very simple calculation. He first observed that a protein consisting of 150 amino acid residues has about 450 degrees of freedom, of which 150 are due to possible variation in

bond angles of the side chains while 300 are due to rotations of peptide bonds (the chemical bonds that keep a chain of amino acids together to form a protein). Assuming that each peptide bond can assume 10 different states (a conservative assumption), there will already be 10^{300} different protein configurations. But this would mean that it takes far too long for a protein to find its state of lowest energy by a random process. For many proteins fold into their correct conformation within just a few seconds Levinthal calculated that even when a protein tries different conformations extremely rapidly, there wouldn't be time to try out more than about 10^{8}. Therefore, the assumption that proteins fold into their state of lowest energy must be wrong. It is much more likely that it finds some local (as opposed to global) energy minimum in a series of steps or "nucleation points" that are due to local interactions in neighboring amino acids.

It is remarkable that a simple consideration such as Levinthal's can topple a widely held theoretical assumption based on an established physical theory (thermodynamics). The case is similar to Crick's: Levinthal also imagined a fictive scenario that defines a space of logical possibilities. Then, he showed that only a small region of this logical space is actually accessible for real proteins to occupy. This, we suggest, is the hallmark of a thought experiment in biology.

5. Artificial Life and Computational Modeling in Biology

Several authors have suggested that artificial life and other computational approaches are a form of thought experimenting (Dennett, 1994; Swan 2009), therefore it is appropriate to briefly discuss these approaches here.

Traditionally, simulation is used to calculate possible fates of dynamical systems for which there is no analytic expression of trajectory. This is often the case in systems described by coupled differential equations, but also in so-called agent-based models,

where a number of agents behave according to simple rules, but collectively produce complex emerging patterns. By revealing these patterns, simulation allows scientists to evaluate whether a particular model possibly explains a phenomenon (Winsberg 2014).

Agent-based models are much used in artificial life (A-life). An example is the virtual world of Tierra created by the evolutionary biologist T. Ray (1993), in which the supposed conditions for biological evolution are reproduced *in silico* to explore the consequences of current theoretical assumptions. More specific studies aim at simulating population behaviors such as the grouping of bird flocks, mammal herds or fish schools, by implementing individual behavioral rules and testing their effect at the population level. These studies suggest possible explanations and proved to be particularly useful in showing what minimal abilities are required from individual agents in order to achieve the complex patterns observed at population level (see Reynolds, 1987, cited in Swan, 2009).

A philosophical analysis of A-life simulations as a kind of thought experimenting has been proposed by Swan (2009), who takes up Daniel Dennett's idea of considering A-life as a way of constructing "thought experiments of indefinite complexity" (Dennett 1994). The starting point of her account is that both in simulations as well as in classical thought experiments the systems under study can be manipulated at will. Specifically, it is possible to examine how a system obeying certain rules will behave. This feature provides a mean to "reason from effects back to probable causes" (Swan 2009, p. 696), which is also the hallmark of abduction as described by Charles Sanders Peirce (1958).[4]

[4] Abduction as characterized by Peirce is the process by which one constructs hypotheses to explain a particular fact. Today, it is sometimes referred to as "inference to the best explanation" (Lipton 2004).

Although these computer simulations seem to perfectly fit the role suggested above – providing hypothetical scenarios to explore the logical field defined by a theoretical framework – they don't satisfy the requirement that thought experiments must involve the imagination.[5] Thus, it seems that in the current literature on thought experiments, the term is used at least in two different senses. In one, wider sense, the label "thought experiment" is given to any kind of theoretical model (including computational models) that involve hypothetical or counterfactual scenarios. In the other, more narrow sense, thought experiments also involve the mental powers of imagination.

6. A Constructive Role for Thought Experiments?

A common feature of all the cases presented here is the exploration of a field of possibilities followed by an evaluation of statements about what is *biologically* possible. The relevant possibilities can include logical possibilities, as in Fisher's idea of evaluating hypothetical three-sex organisms in order to understand why there mostly are only two. Alternatively, the salient possibilities may be constrained by the principles of a theoretical framework, like in the case of Crick's discussion of the genetic code, in Levinthal's paradox, life history traits, or evolutionary simulations of the A-life application Tierra. The outcome of the thought experiment is an evaluation of statements expressing such logical or theoretical possibilities, which seems to depend on the accordance with further biological principles pertaining to the particular region of the logical space that is visited. In general, it seems that antecedent theoretical knowledge plays a major role in most cases, so biological thought experiments for the most part don't just rely on untutored intuitions.

[5] This is the reason why Chandrasekharan, Nersessian and Subramanian (2012) don't consider computer simulations as thought experiments but rather as an alternative approach.

In some cases, thought experiments reveal constraints that had not yet been taken into account in the theoretical framework. Examples for this include the time constraints in Levinthal's paradox, the impossibility of an infinity of specific biosynthetic enzymes in Crick's reasoning about the mechanism of protein synthesis or the inexistence of a Darwinian demon. When the general theoretical framework cannot cope with the particular constraints involved in the hypothetical case, thought experimenting seems to serve the purpose of pinpointing explanatory defects or eventually the existence of a problematic assumption in the theoretical framework. In other cases, the thought experiment seems to support the idea that the proposed theoretical framework sufficiently explains typical hypothetical cases, like in Darwin's thought experiments or some A-life simulations.

Some of the experiments discussed above seem to occur in a context of discovery in a wider sense, where novel hypotheses are produced and evaluated prior to experimental verification. Since thought experiments carry some evidential power in the assessment of possible explanatory hypotheses, but seem nevertheless unable to provide sufficient confirmation, their role might be construed along the line of Curd's (1980) "assessment" of research hypotheses, which occurs in the context of discovery rather than the context of justification because it cannot provide sufficient justification, but has some justificatory value since it allows to make an informed choice among a set of possible hypotheses.

In other cases, thought experiments might not lead directly to testable hypotheses, like in the case of some A-life simulations, Levinthal's paradox, or Darwin's demon, but their role there is to reveal a lack of explanatory constraints and the need for further theoretical hypotheses.

This picture of thought experimenting as exploring and evaluating a field of theoretical possibilities suggests two kinds of theorizing activities that relate to the imagination criterion. The first activity is the choice of theoretical assumptions that should constitute the most

relevant constraints pertaining to the particular situation. The second activity consists in inferring the consequences following from these constraints. These consequences seem mainly to be obtained either by deduction or by computational simulation, but other means have been proposed, like the appeal to mental models suggested by Nersessian (1992).[6] Since in simulations the step of inferring the consequences of a set of assumptions is delegated to a computer, their status as thought experiment may be denied (Chandrasekharan, Nersessian and Subramanian 2012; cf. also the entry on computer simulations in this book), but the choice of the assumptions still depends on the theoretician's imagination and this may suffice to view, with Dennett and Swan, computational simulations as extended thought experiments, just like the use of paper and pencil suitably enhances geometrical thought experiments. Our examples thus seem to provide a continuum of cases where some rely more on the evaluation of theoretical assumptions, like in the Darwinian thought experiments, whereas others highlight more the inferential aspects, be it deductive or computational, like in the case of Levinthal's paradox or A-life simulations. But it seems that both assumption evaluation and inferential work are present as aspects in all the cases.

The examples presented in this article suggest that thought experimenting may widely be used in scientific modeling as a mean of evaluating the relevance of a theoretical model as well as its consistency with other models pertaining to similar cases. In this picture, thought experiments are less important than sometimes assumed for the justification of particular statements, but they appear to be much wider spread in the scientific practice. They may play

[6] It is not clear in the case of the view defending Platonic insights defended by Brown (1991), wether these insights occur at the level of the choice of assumptions, inference of their consequences or both. Maybe though it could be considered as another way of inference making to be added to this list.

an important role for the integration of theoretically scattered scientific fields and provide guidance for scientific research.

The results provided here should be relativized in any case to the selection of cases we made in the first place. Our broad criteria allowed us to choose what we consider the most significant cases of thought experimenting in biology, but philosophers assuming a more specific view of thought experimenting may end up with a different picture.

References

Bonsall, Michael B. (2006), Longevity and Ageing: Appraising the Evolutionary Consequences of Growing Old, *Philosophical Transactions of the Royal Society of London B* 361:119-135.

Brown, J. R. (1991). *The laboratory of the mind: Thought experiments in the natural sciences*. London: Routledge.

Bulmer, M. (1999). Did Jenkin's swamping argument invalidate Darwin's theory of natural selection? *The British Journal for the History of Science, 37*(3), 281–297. doi:10.1017/S0007087404005850

Chandrasekharan, S., Nersessian, N. J., & Subramanian, V. (2012). Computational Modeling: Is This the End of Thought Experiments in Science. In : *Routledge studies in the philosophy of science, Thought experiments in science, philosophy, and the arts*, pp. 239–260. London: Routledge.

Crick, Francis H.C. (1958), On Protein Synthesis, *Symposia of the Society for Experimental Biology* 12:138-163.

Curd, M. V. (1980). The logic of discovery: an analysis of three approaches. In T. Nickles (Ed.), *Scientific discovery, logic, and rationality* (pp. 201–219).

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. or the Preservation of Favoured Races in the Struggle for Life. London: John Murray.

Darwin, C. (1871). *The Descent of Man*. And Selection in Relation to Sex. New York: D. Appleton and Company.

Dennett, D. (1994). Artificial life as philosophy. *Artificial Life, 1*(3), 291–292.

Fisher, R. A. (1930), *The Genetical Theory of Natural Selection*. Oxford: Clarendon.

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society B: Biological Sciences, 205*(1161), 581–598. doi:10.1098/rspb.1979.0086

Hull, D. L. (1973). *Darwin and his critics. The reception of Darwin's theory of evolution by the scientific community* (Vol. 1).

Judson, Horace Freeland (1979), *The Eighth Day of Creation. Makers of the Revolution in Biology*. New York: Simon and Schuster.

Kitcher, P. (1982). *Abusing science: The case against creationism*. Cambridge, Mass: MIT Press.

Kitcher, P. (1985). *Vaulting ambition: Sociobiology and the quest for human nature*: Mit Press Cambridge, MA.

Lange, Marc, and Alexander Rosenberg (2011), Can There Be A Priori Causal Models of Natural Selection?, *Australasian Journal of Philosophy* 89 (4):591-599.

Law, Richard (1979), Optimal Life Histories Under Age-specific Predation, *The American Naturalist* 114 (3):399-417.

Lennox, J. G. (1991). Darwinian thought experiments: A function for just-so stories. In T. Horowitz & G. Massey (Eds.), *Thought experiments in science and philosophy* (pp. 223–245).

Levinthal, Cyrus (1969), How to Fold Graciously, in J.T.P. DeBrunner and E. Munck (eds.), *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois*, Champaign-Urbana: University of Illinois Press.

Lipton, Peter (2004), *Inference to the Best Explanation, 2nd Edition*. London: Routledge.

Mayr, Ernst (1982), *The Growth of Biological Thought*. Cambridge Mass.: Harvard University Press.

Morris, S. W. (1994). Fleeming Jenkin and The Origin of Species: a reassessment. *The British Journal for the History of Science, 27*(03), 313. doi:10.1017/S0007087400032209

Nersessian, N. J. (1992). In the theoretician's laboratory: Thought experimenting as mental modeling. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (pp. 291–301).

Peirce, C. S., & Burks, A. W. (1958). *Collected Papers of Charles Sanders Peirce: Reviews, Correspondence, and Bibliography; Edited by Arthur W. Burks*: Harvard University Press.

Provine, William B. (1971), *The Origins of Theoretical Population Genetics*. Chicago: University of Chicago Press.

Ray, T. S. (1993). An evolutionary approach to synthetic biology: Zen and the art of creating life. *Artificial Life, 1*(1_2), 179–209.

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Computer Graphics, 21*(4), 25–34.

Sober, E. (2011). A Priori Causal Models of Natural Selection. *Australasian Journal of Philosophy, 89*(4), 571–589. doi:10.1080/00048402.2010.535006

Stearns, S. C. (1992), *The Evolution of Life Histories*. Oxford: Oxford University Press.

Swan, L. S. (2009). Synthesizing insight: artificial life as thought experimentation in biology. *Biology & Philosophy, 24*(5), 687–701.

Waters, C. K. (2003). The arguments in the *Origin of Species.* In J. Hodges & G. Radick (Eds.), *The Cambridge Companion to Darwin* (pp. 116–139). Cambridge University Press.

Weisberg, Michael (2013), *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

Winsberg, Eric (2014), "Computer Simulations in Science", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2014 edition)*.